社会化阅读环境下阅读内容计量研究*

——以阅读推广类微信公众号推送文章为例

王磊[1]　江仔玲[2]　刘宇婷[3]

[1] 合肥工业大学图书馆　合肥 230009　[2] 合肥工业大学仪器科学与光电工程学院　合肥 230009

[3] 苏州大学纺织与服装工程学院　苏州 215021

摘要：[**目的/意义**] 随着社交媒体大数据平台的构建、大数据挖掘、分析技术及数据库技术的不断演进，社交媒体中各种数据变得可获取、可利用。据此，开展了基于社交媒体数据的社会化阅读内容计量研究。[**方法/过程**] 首先，利用大数据平台及文本挖掘技术对阅读推广类微信公众号基本信息和公众号推送文章内容进行采集。其次，对采集的数据进行统计，得到了各微信公众号在分类、地域、认证上的分布，以及各微信公众号推送文章数、阅读数、平均阅读数、点赞数、平均点赞数，并通过运用信息计量学相关理论及技术，利用自编脚本得到了单篇推文传播指数(Single Tweets Communication Index, STCI)、标题高频关键词共现矩阵、基于推文重复次数的微信公众号耦合矩阵。同时，提出最高 STCI 微信推文发布延时指数(Publish Delay Index of Highest STCI Wechat Ariticle, HSPDI)算法来考察推文重复发布次数对推文传播效果的影响。最后，采用人工研读高 STCI 推文的方法对其内容特征进行归纳。[**结果/结论**] 阅读推广类微信公众号推送文章的平均阅读数与平均点赞数存在一定的正相关关系；许多微信公众号之间存在严重的推送相同文章现象；文章是否是微信公众号首次推送对其传播没有明显影响。进一步，对高 STCI 推文内容分析后得出高 STCI 推文特征主要包括：满足读者的好奇心、满足读者的自我表达需要和满足"屏阅读"时代特征的写作法则。

关键词：社会化阅读　社交媒体　微信公众号　信息计量

分类号：G252

DOI：10.13266/j.issn.0252-3116.2019.02.002

1 引言

古往今来，阅读一直被认为是推动社会文明程度提高的重要途径[1]。而阅读推广在传承文化、开发智源、促进创新和提振人类精神力量等方面发挥着巨大作用[2]。纵观人类阅读的发展史，在仅有纸介媒体的时代里阅读活动是线性的[3]。在此期间，读者间虽然存在互动和分享，但读者的互动和分享行为往往被局限于很小的物理空间，例如传统的读书会[4]。随着第三代科技革命兴起，互联网络技术被大力发展，人与技术的关系变得密不可分，阅读内容的载体也从纸质文献扩展到了电子媒介，阅读也由纸本阅读延伸到了数字阅读[5]。而后 web2.0 时代的到来，传统知识信息单向传播、大众传播的特性被打破，双向乃至多向的知识交互传播和分众传播得以实现[6]。网络用户迈入了更为主动的时代，他们不仅是信息的接受者，还掌握着信息生产和传播的主动权，扮演着信息的创造者和传播者的角色[7]。此背景下，基于分享与互动的社交媒体阅读已然成为大众青睐的社会化阅读方式，其中最具代表性的社交媒体平台有美国的 Twitter 和 Facebook，中国的微信、微博等[8]。在社交媒体上深入探索读者阅读行为和阅读习惯的社交化转型，以期提供更便捷、更人性化的社会化阅读服务已成为当前业界和学术界共同关注的热点问题，中外学者们从不同的角度对社会化阅读进行了大量的研究，研究内容主要包括社会化阅读中用户阅读行为分析[9-10]、阅读信息的传播与

* 本文系中央高校基本科研业务费专项资金资助项目"扎根社交媒体大数据的社会化阅读探究"(项目编号：JS2018HGXJ0042)研究成果之一。

作者简介：王磊(ORCID：0000-0002-8000-7223)，信息咨询与学科服务部主管，副研究馆员，博士，E-mail：wangl@hfut.edu.cn；江仔玲(ORCID：0000-0002-3843-8740)，本科生；刘宇婷(ORCID：0000-0003-2597-3944)，本科生。

收稿日期：2018-07-03　修回日期：2018-10-12　本文起止页码：12-23　本文责任编辑：杜杏叶

使用[11-14]、社会化阅读对青少年阅读的影响[15-18]及基于社会化阅读的图书馆阅读推广策略[19-22]。

然而,本文发现目前研究缺少从文本内容维度考量社会化阅读,因此开展了基于社交媒体数据的社会化阅读内容计量研究可行性探索。在社交媒体数据选择中,本文选择了阅读推广类微信公众号基本数据和微信公众号推送文章(微信推文)内外部特征数据。研究主要涉及如何实现微信公众号数据采集;如何实现微信公众号传播能力与单篇推文传播能力计量(STCI);如何实现基于推文重复次数的微信公众号相似性计量;如何设计并实现重复推文中最高STCI推文的延时计量;如何实现高STCI推文内容分析并提取推文特征。希望通过本文可以为学界同仁揭示国内阅读推广类微信公众号的基本特征以及提供一种可借鉴的基于社交媒体数据的社会化阅读内容计量新视角,使得图书情报学中社交媒体大数据的应用突破Altmetrics研究范畴,让新时代图书馆社会化阅读理论与方法研究也能从社交媒体大数据的发展中获益。

2 研究方法

本文选择阅读推广类微信公众号作为研究对象。其主要原因如下:①相对于其他的专业社会化阅读APP如Zite、Flipboard、网易云阅读和扎客,微信阅读的用户数量更为庞大。如《2017微信数据报告》中指出同年9月份,每天平均有9.02亿人登陆微信。②阅读推广类微信公众号的推文通常是具有一定篇幅的长文本,因此阅读这样的微信推文是真正意义上的"阅读行为"。③当用户通过微信公众号阅读推文(特别是阅读由好友分享的推文)时,社交行为本身就产生了阅读入口。且订阅微信公众号虽属于个人行为,但用户在阅读过程中经常开展分享或互动活动,这非常符合社会化阅读的核心特征[23]。同时,通过调查本文选择中国新媒体大数据评价体系和影响力标准的研究制定者——清博指数大数据平台作为研究对象的数据源[24]。数据源采集及数据分析流程如图1所示:

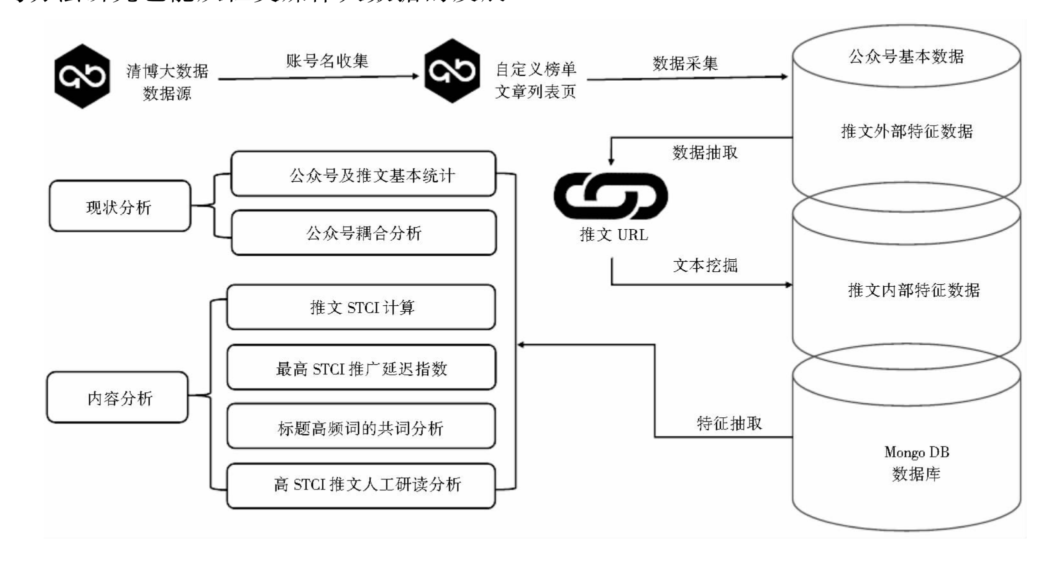


图1 微信公众号数据源采集及数据分析流程

根据上图1,本文的设计流程可以分为数据采集与数据分析两个部分。本文首先利用清博大数据平台导出所研究微信公众号的基本数据(微信号、账号名、分类、认证信息、地区、功能介绍、标签)和微信推文外部特征数据(文章标题、阅读数、点赞数、url);通过外部特征中的url链接,使用文本挖掘技术采集各微信公众号的微信推文内部特征数据(推文内容、推文内容字数、动态图片数、静态图片数、图片数、音频数、视频数、标题字数),并将这些数据存储在mongo DB数据库里。之后,利用Python语言自编脚本得到了单篇推文传播指数(Single Tweets Communication Index, STCI)、标题高频关键词共现矩阵、基于推文重复次数的微信公众号耦合矩阵。同时,提出最高STCI推文发布延时指数(Publish delay index of highest STCI Wechat ariticle, HSPDI)算法来考察推文发布时间先后对其传播效果的影响。最后采用人工研读高STCI推文的方法对微信公众号推文的内容特征进行归纳。虽然本次研究的数据收集量及分析量不算太大(56 355条结构化数据),但本文所编写的程序脚本和便于拓展的mongo DB数据库完全可用于大数据分析。

2.1 数据源采集

首先在清博大数据平台中搜索"tag:阅读推广",

共得到 395 个相关结果，基于“二八定律”（事物的关键部分起决定作用），选择其中 WCI 指数排名前20% 的 79 个进行研究，将其添加到自定义榜单。表 1 为阅读推广类微信公众号、账号名和 WCI 指数的统计结果。

之后，根据 79 个阅读推广类微信公众号的基本信息，利用清博大数据平台的自定义榜单导出功能将 2017 年 5 月 1 日至 2017 年 10 月 31 日期间内 56 355 篇微信推文的外部特征（文章标题、阅读数、点赞数、url）导出。利用文本挖掘技术遍历这 56 355 篇微信推文中所有 url 链接，得到对应推文的内部特征数据，整合后存储在 Mongo DB 数据库中。表 2 为存储在 Mongo DB 数据库的部分数据。

表 1　79 个阅读推广类微信公众号微信号、帐号名和 WCI 指数

微信号	账号名	WCI	微信号	账号名	WCI	微信号	账号名	WCI
人卫中医	Renweizhongyi	433.16	新京报书评周刊	Ibookreview	861.62	为你读诗	Thepoemforyou	1 086.94
书橱阅读	shuchu_com	673.15	有故事的人	ifengstory	872.31	有书共读	youshu_cc	1 087.73
未读	unreadsky	710.07	文刊	aabb936	874.85	言情说	yqsaying	1 088.62
百草园精选	bcyjingxuan	723.18	365 读书	dus365	878.73	书单	BookSelection	1 098.38
美文共享	MeiwenGX	723.47	围炉夜读	weiluyedu_	879.02	格上阅读	GSyuedu	109.62
每天一读	mtyidu	724.02	本真图书	benzhenbook	880.23	壹读	yiduiread	1 127.4
地图帝	MapEmperor	735.06	读首诗再睡觉	dushoushizaishuijiao	884.5	微信读书	weread_tencent	1 155.08
叶之秋读书	yezhiqiu1979	736.79	凤凰读书	ifengbook	885.46	读书 369	dushu369com	1 187.62
短篇集	duanpianji	738.06	拾读文摘	sdds10	885.54	品读时刻	pindushike	1 202.05
笨鸟文摘	bnwz2012	751.15	论道	lundao11	886.2	悦读	yuedu58	1 217.2
Kindle 电子书库	kindle10000	768.53	大象公会	idxgh2013	890.54	一读	Iiiread	1 248.24
法律读库	lawreaders	772.82	疯狂阅读	fkydzz	899.8	书单来了	shudanlaile	1 249.6
花边阅读	huabianyuedu	772.98	趣阅	quyue333	901.98	十点文摘	shidian3650	1 260.15

表 2　Mongo DB 数据库中数据（部分）

公众号	采集日期	题目	阅读数	点赞数	静态图片	动态图片	音乐数	视频数	采集时间	STCI	文章字数
国学文化	2017-06-01	致 童 年！	100 001	9 015	0	1	0	0	159	1 093.7	447
读文摘精选	2017-06-01	童年钟声	100 001	4 157	2	1	0	0	159	1 029.93	942
格上阅读	2017-06-01	【豪送现金】518 元现金红包，人人有份！	100 001	699	0	0	0	0	159	890.55	0
大象公会	2017-06-01	红领巾是怎样染红的丨大象公会	100 001	619	0	0	0	0	159	881.45	0
十点读书	2017-06-01	永远不要小看不合群的人	100 001	30 211	4	3	1	0	159	1 197.19	1 480
有书	2017-06-01	婚姻好不好，看厨房就知道	100 001	15 454	3	10	1	0	159	1 139.25	1 920
十点读书	2017-06-01	一开始喜欢你的人，后来为什么不喜欢了？	100 001	10 064	6	4	1	0	159	1 102.93	2 158
百草园书店	2017-06-01	早读：愿你我的纯真不变，初心依旧	100 001	2 705	11	1	1	0	159	995.37	813
十点读书	2017-06-01	没有一种工作是不委屈的	100 001	14 646	7	4	1	0	159	1 134.67	2 580

2.2　数据分析

2.2.1　微信公众号基本数据统计及基于推文重复次数的微信公众号耦合分析　通过运行自编的程序脚本，得到采集日期内微信公众号的推文数、阅读数、平均阅读数、点赞数、平均点赞数以及各微信公众号在分类、认证信息、地域分布上的数据统计结果。

在公众号耦合分析中，以各公众号之间推文的重复次数来定义各公众号之间的耦合关系强度。首先，利用自编程序脚本获得微信公众号在采集日期内重复推送推文（共 5 983 篇）的基本信息。之后，针对每个公众号统计其与其他各公众号的推文重复次数，选择该重复次数为相应两个公众号之间的耦合强度。通过自编算法得到耦合矩阵，并利用 NetDraw 软件将矩阵数据可视化，得到基于重复次数的各公众号耦合网络谱图。

2.2.2　微信传播指数 WCI 与单篇推文传播指数 STCI 指标算法介绍　清博指数大数据平台中的 WCI（微信传播指数 V13.0）[25] 比之前的版本有了更丰富的指标，从而得到了更全面的评价维度，且优化了各项指标的权重，使之所得排名更加精确，其计算如公式（1）所示：

$$WCI = \{30\% * [0.85 * \ln(R/d + 1) + 0.15 * \ln(10 * Z/d + 1)] + 30\% * [0.85 * \ln(R/n + 1) + 0.15 * \ln(10 * Z/n + 1)] + 30\% * [0.85 * \ln(Rt/d + 1) +$$

$0.15 * \ln(10 * Zt/d + 1)] + 10\% * [0.85 * \ln(Rmax + 1) + 0.15 * \ln(10 * Zmax + 1)]\}^2$ （公式 1）

其中，n 为数据采集日期内公众号推送文章数；R 为数据采集日期内所有推文(n)的阅读数；Z 为数据采集日期内所有推文(n)的点赞数；d 为数据采集日期所含天数；Rt 和 Zt 为数据采集日期内公众号所推送头条的阅读总数和点赞总数；Rmax 和 Zmax 为数据采集日期内公众号推送文章的最大阅读数和最大点赞数。但该指数主要用于评估微信号的整体热度和发展走势，而本文需要找到高传播指数的单篇推文，为此本文利用前期的研究成果单篇推文传播指数 STCI(Single Tweets Communication Index)[26] 来评估单篇推文传播效果，其计算如公式(2)所示：

$$STCI = \{60\% * (20\% * \ln(R/d + 1) + 80\% * \ln(R + 1)) + 40\% * (20\% * \ln(10 * Z/d + 1) + 80\% * \ln(10 * Z + 1))\}^2 * 10 \quad (公式 2)$$

其中，R 为当篇推文自推送到数据采集截止日内的推文阅读量；Z 为当篇推文自推送到数据采集截止日内的推文点赞量；d 为当篇推文自推送到数据采集截止日内所含天数。本文数据收集时间为 2017 年 11 月 07 日，采集数据时间段为 2017 年 5 月 1 日至 2017 年 10 月 31 日(如果当篇推文的推送日期为 2017 年 5 月 1 日，那么 d = 190 天)。

2.2.3 最高 STCI 推文延迟指数算法介绍　为得出被重复推送的推文推送时间先后与 STCI 指数的关系，提出最高 STCI 推文延迟指数(Publish Delay Index of highest STCI Wechat Ariticle, HSPDI)，具体计算步骤为：①从推文外部特征数据中抽取出同一推文的所有推送日期 $\{date_2, date_5, date_1, date_4, date_3 \cdots date_n \cdots date_i\}$ 及相应推文的 STCI 指数 $\{STCI_2, STCI_5, STCI_1, STCI_4, STCI_3 \cdots STCI_n \cdots STCI_i\}$；②将每个日期集合与相应的 STCI 指数集合按时间先后顺序进行排序得到同一推文的发布日期排序集合 $\{date_1, date_2, date_3, date_4, date_5 \cdots date_i \cdots date_n\}$ 和对应的单篇推文 STCI 指数集合 $\{STCI_1, STCI_2, STCI_3, STCI_4, STCI_5 \cdots STCI_i \cdots STCI_n\}$；③比较得到同一推文的最高 STCI 指数 $STCI_i$，并采集其排列序数 i；④将排列序数 i 与集合个数 n 相除得到推文的最高 STCI 推文延迟指数。其计算模型如下：

$$DATE = \{date_2, date_5, date_1, date_4, date_3 \cdots date_n \cdots date_i\}$$
$$STCI = \{STCI_2, STCI_5, STCI_1, STCI_4, STCI_3 \cdots STCI_n \cdots STCI_i\}$$
$$\downarrow 排序$$
$$DATE' = \{date_1, date_2, date_3, date_4, date_5 \cdots date_i \cdots date_n\}$$
$$STCI' = \{STCI_1, STCI_2, STCI_3, STCI_4, STCI_5 \cdots STCI_i \cdots STCI_n\}$$
$$Max(STCI') = STCI_i \rightarrow 取\ i \rightarrow HSPDI = i/n$$

其中集合 DATE 为同一推文的不同发布日期集合，集合 STCI 为对应的推文 STCI 指数集合，集合 DATE' 为排序后同一推文的发布日期集合，集合 STCI' 为排序后对应的单篇推文 STCI 指数集合，n 为集合中的元素个数。

具体计算实例：推文《致所有不读书的人：你将“死”于 35 岁》的重复次数为 n = 4，其所有发布日期为 DATE = [2017-09-02, 2017-09-28, 2017-08-23, 2017-09-03]，相应的 STCI 指数为 STCI = [658.54, 425.8, 416.09, 586.47]。按日期排序后变为 DATE' = [2017-08-23, 2017-09-02, 2017-09-03, 2017-09-28]，相应的 STCI 指数为 STCI' = [586.47, 658.54, 416.09, 425.8]。比较得到最高的 STCI 指数为 $STCI_2$ = 658.54，排列序数为 i = 2，故由 HSPDI 计算公式得其最高 STCI 推文延迟指数为 i/n = 2/4 = 0.5。

2.2.4 微信推文内容分析及基于标题高频词的共现分析　在基于标题分词的共现分析中，从 mongo DB 数据库中抽取每条结构化数据的推文标题，并通过“结巴”中文分词模块得到标题具有实际意义的名词(n)、动词(v)、动名词(vn)。通过自编的程序脚本，统计出标题高频词和高频词共现矩阵。最后，将数据导入 Ucinet 软件并利用 NetDraw 可视化软件绘制推文标题高频词共现网络图谱。为进一步对微信公众号推文内容分析，本文通过人工逐篇阅读高 STCI 推文和低 STCI 推文的方法对微信公众号推送文章的内容特征进行归纳。

3 结果与讨论

3.1 阅读推广类公众号推文整体分析

统计并分析 2017 年 5 月 1 日至 2017 年 10 月 31 日期间内采集到的 56 355 条结构化数据，得到 79 个阅读推广类公众号的推文数、阅读数、平均阅读数、点赞数、平均点赞数，图 2 为可视化后的统计图表。

其中，推文数是指各公众号采集日期内所有推文的总和，阅读数是指各公众号采集日期内所有推文的阅读量总和，平均阅读数是指各公众号阅读数除以推文数，点赞数是指各公众号采集日期内所获总赞数，平均点赞数是指点赞数除以推文数。显而易见，推文数前五名为美文阅读(1 472 篇)、水木文摘(1 471 篇)、百草园书店(1 470 篇)、慈怀读书会(1 470 篇)、读书369(1 467 篇)；阅读数前五名为十点读书(1.46 亿次)、

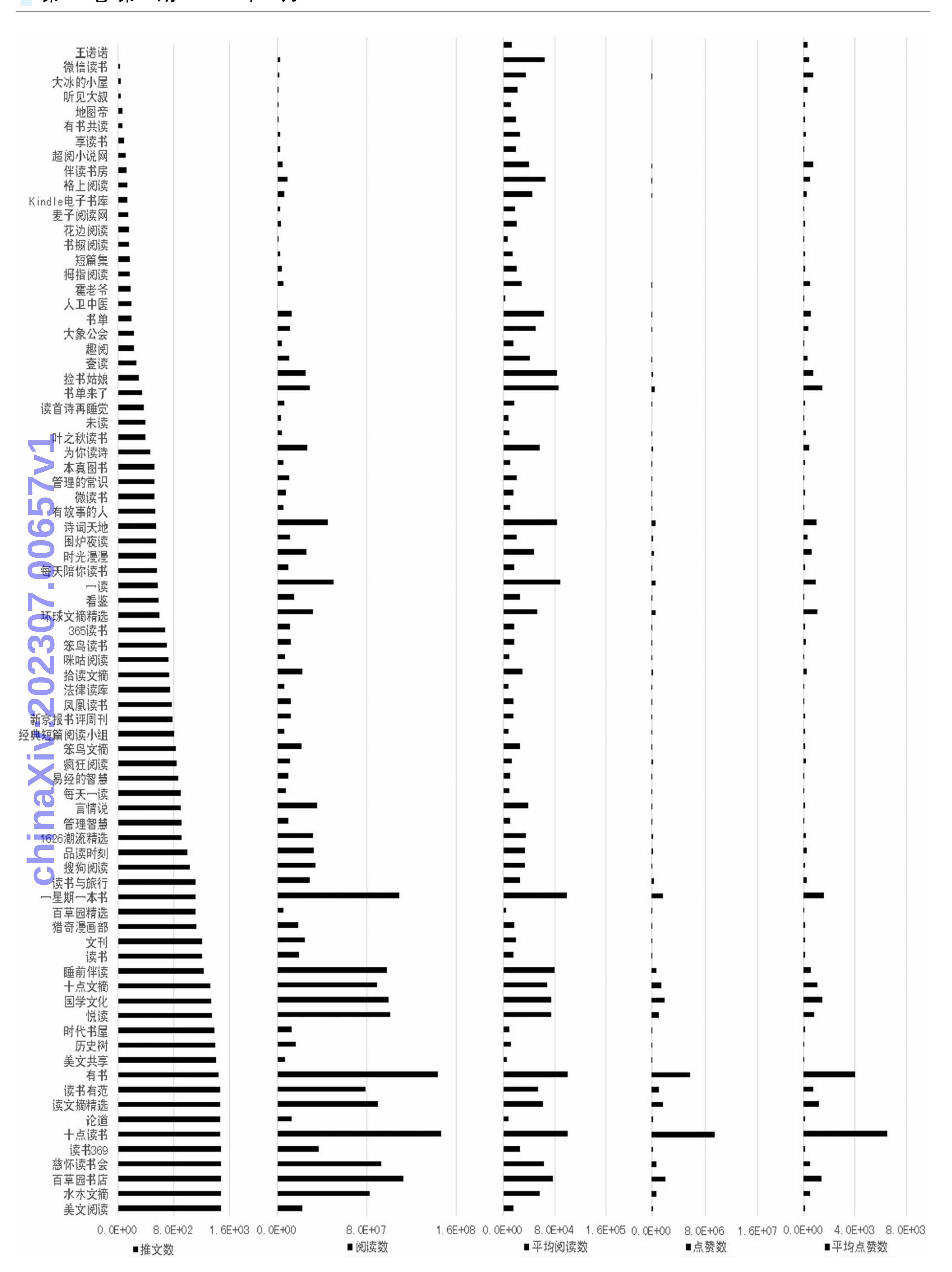


图 2　公众号推文数、阅读数、平均阅读数、点赞数、平均点赞数统计

有书(1.44 亿次)、百草园书店(1.13 亿次)、一星期一本书(1.09 亿次)、悦读(1.01 亿次);平均阅读数前五名为十点读书(10 万 +)、有书(9.99 万次)、一星期一本书(9.91 万次)、一读(8.82 万次)、书单来了(8.56 万次);点赞数前五名为十点读书(949.78 万次)、有书(576.29 万次)、百草园书店(203.57 万次)、国学文化(196.19 万次)、一星期一本书(175.52 万次);平均点赞数前五名为十点读书(6 487 次)、有书(4 007 次)、一星期一本书(1 589 次)、书单来了(1 480 次)、国学文化(1 470 次)。

由图 2 可知公众号推文数与阅读数、平均阅读数、点赞数、平均点赞数均不存在明显的相关关系,而平均阅读数、平均点赞数之间存在较强的正相关关系。换而言之,推文数高,平均阅读数和平均点赞数不一定就高,但平均阅读数高,其平均点赞数往往就高。这在一定程度上说明读者在阅读完文章标题后是否点击进去阅读是由个人主观决定的,但是点击阅读后点赞或不点赞却是根据个人习惯决定的。这也说明了推文标题对推文传播的重要性。

3.2 基于分类、认证、地域分布的公众号统计

从上述 79 个公众号的基本数据中抽取分类、认证信息、地域分布数据,如表 3 所示(部分数据):

表 3　公众号分类、认证信息、地域分布数据统计(部分)

微信号	分类	地区	认证信息	WCI
有书	文化	北京市	北京万维之道信息技术有限公司	1 435.15
十点读书	文化	福建省	厦门十点文化传播有限公司	1 439.76
百草园书店	文化	江苏省	无锡市百草园书店有限公司	1 368.88
诗词天地		－	郑州坚果互娱网络科技有限公司	1 327.58
一星期一本书	文化	江苏省	南京口袋网络科技有限公司	1 359.9
国学文化	文化	上海市	暂未认证	1 278.73
读文摘精选	文化	广东省	深圳市斯必得机电设备有限公司	1 286.42
睡前伴读	生活	江苏省	南京初美文化传媒有限公司	1 285.22
十点文摘	文化	江苏省	徐州雨轩文化传媒有限公司	1 260.15
读书有范	文化	上海市	上海初色网络科技有限公司	1 270.52

本文将 79 个阅读推广类公众号的来源地进行统计,发现北京市拥有的微信公众号最多,有 15 个,其次是广东省 9 个、江苏省 6 个,上海、山东、河南也相对较多,分别为 4 个、3 个、3 个。从区域分布状况来看,微信公众号主要集中分布在沿海及华中经济发达地带。结合图 2 可以看出,阅读数前 5 名的十点读书、有书、百草园书店、一星期一本书、悦读,点赞数前 5 名的十点读书、有书、百草园书店、国学文化、一星期一本书中大部分为认证公众号,说明获得认证的公众号更具有公信力,读者更倾向于选择获得认证的微信公众号推文进行阅读。结合表 3 的 WCI 指数排名和公众号是否认证可以看出,高 WCI 指数的公众号大部分已获得认证,由此本文得出结论,微信公众号是否认证与其推文社会化阅读推广效果有关,即微信公众号是否认证会影响公众号社会化阅读推广效果。

3.3 基于重复推文次数的各公众号耦合分析

通过自编算法,本文得到基于重复推文次数的 79 个微信公众号耦合矩阵,具体数值见表 4(部分数据)。这是一个网络密度为 3.41 的全联通网络,由于网络中连线较密,所以很难在可视化后对网络细节进行描述。为了更清晰地体现基于相似性的各公众号耦合网络中各公众号之间的关系,本文抽取耦合强度值大于等于 30 的数据进行可视化,去除掉孤立节点后得到了深度聚合的公众号关系网络,见图 3。图中圆点表征公众号,圆点大小表征该公众号所发布的重复推文数,圆点越大,所发布的重复推文越多。圆点间连线的粗细表征所连两个公众号之间的耦合强度大小。

虽然阅读推广类公众号之间没有联合推文,但是利用推文的重复次数来定义各公众号之间的耦合关系将会揭示其间的潜在关系。换而言之,这些公众号也许没有有组织的联合推文,但却在自身推文体系中关注了相同内容的推文。从图 3 可知,读书与悦读、一星期一本书与睡前伴读、悦读与水木文摘、国学文化与读文摘精选、读文摘精选与十点文摘之间的耦合强度为所有耦合强度的前五位,分别为 429、396、374、359、341。结合表 1,读书的 WCI 指数为 908.03,而悦读的 WCI 指数五 1 217.2,两者之间差距较大,显示耦合强度大的两个公众号之间 WCI 指数也可能相差较大。但一星期一本书的 WCI 指数为 1 359.9,睡前伴读的 WCI 指数为 1 285.22;国学文化的 WCI 指数为 1 278.73,读文摘精选的 WCI 指数为 1 286.22;这又显示出耦合强度大的公众号 WCI 指数相差不大。因此,本文认为阅读推广类微信公众号之间互相借鉴推文是普遍现象,拥有相同推文较多(关系较强)的公众号之间传播能力不尽相同。

3.4 最高 STCI 推文延迟指数(HSPDI)分析

利用自编程序脚本提取出各微信公众号中所有被重复推送过的推文(共 5 983 篇)的外部特征数据。通过运行 HSPDI 算法脚本,得到这 5 983 篇推文的最高 STCI 推文延迟指数(HSPDI)。

基于日常生活中新闻独家性与时效性观念,本文猜想发布时间越早的推文的 STCI 指数越高,即最高 STCI 推文延迟指数应为重复次数的倒数。为验证此

表4 基于相似性的各公众号耦合矩阵

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	400	9	0	0	0	0	36	27	41	9	16	0	20	11
B	9	201	0	0	0	0	15	10	5	0	3	0	1	2
C	0	0	17	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	7	0	0	0	0	1	0	0	0	0	0
E	0	0	0	0	1	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	27	0	1	1	0	0	0	0	0
G	36	15	0	0	0	0	524	7	6	5	24	0	69	9
H	27	10	0	0	0	1	7	312	74	12	4	0	6	2
I	41	5	0	1	0	1	6	74	790	28	6	0	9	2
J	9	0	0	0	0	0	5	12	28	156	2	0	3	3
K	16	3	0	0	0	0	24	4	6	2	131	0	14	0
L	0	0	0	0	0	0	0	0	0	0	0	38	0	0
M	20	1	0	0	0	0	69	6	9	3	14	0	264	3
N	11	2	0	0	0	0	9	2	2	3	0	0	3	117

注:A:笨鸟文摘;B:百草园精选;C:读首诗再睡觉;D:叶之秋读书;E:大冰的小屋;F:本真图书;G:十点读书;H:品读时刻;I:十点文摘;J:每天陪你读书;K:围炉夜读;L:有故事的人;M:疯狂阅读;N:管理的常识

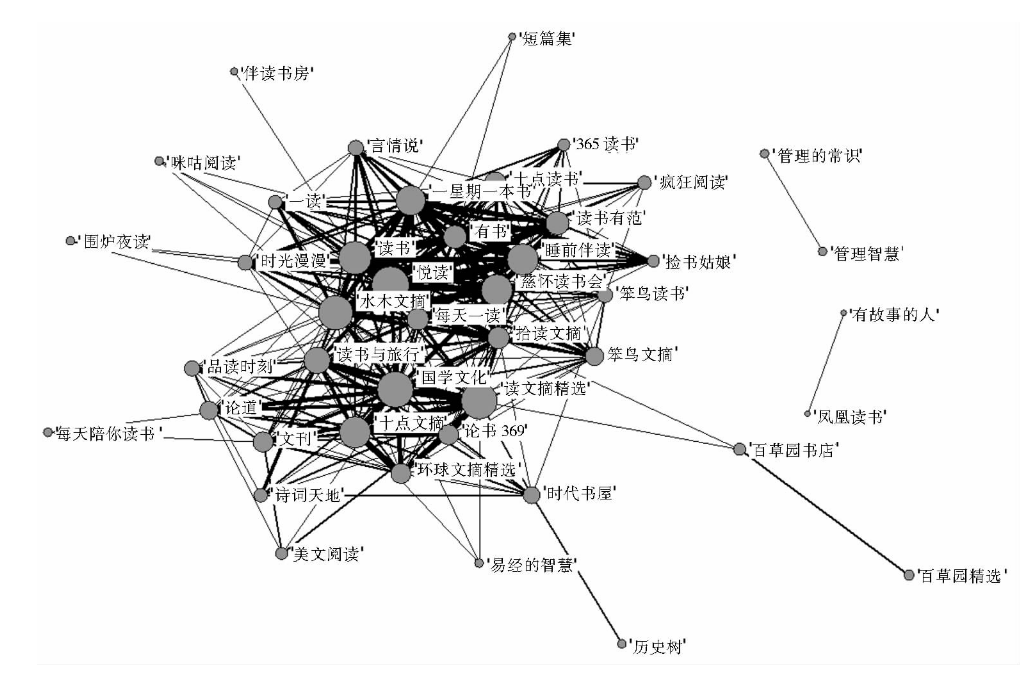


图3 基于相似性的公众号耦合网络可视化图谱

猜想,本文通过运行自编程序脚本,得到每篇推文(5 983篇)的最高STCI推文延迟指数,并以推文重复数作为横坐标,最高STCI推文延迟指数作为纵坐标,利用Matplotlib库进行可视化,得到分布图4。每个十字光标对应一篇推文,其明亮程度代表重复次数相同的情况下HSPDI指数相同的推文篇数的多少,灰色圆点代表重复次数相同推文的最小HSPDI指数(即重复次数的倒数)。显然,每篇推文对应的十字光标及明亮程度呈随机分布,其HSPDI指数并不是本文所设想的为最小HSPDI值。另外,在分析推文的重复次数及其STCI指数的关系时,本文发现重复次数排名前三的推文《花钱,最能看出一个人的基本素质和心态》重复35次、《一个女人最好的生活状态》重复30次、《女人过得好不好,从一点就能看出来》重复23次,它们的最高STCI指数分别为777.38、1 164.91、758.98。

由此本文推断推文发布时间的先后与其STCI指数无关,这在一定程度上说明用户在阅读微信推文时并不关心其是否为原创。同时,由重复次数排名前三的推文STCI指数与高STCI推文的STCI指数比较可以看出推文重复次数与推文的传播能力无关。

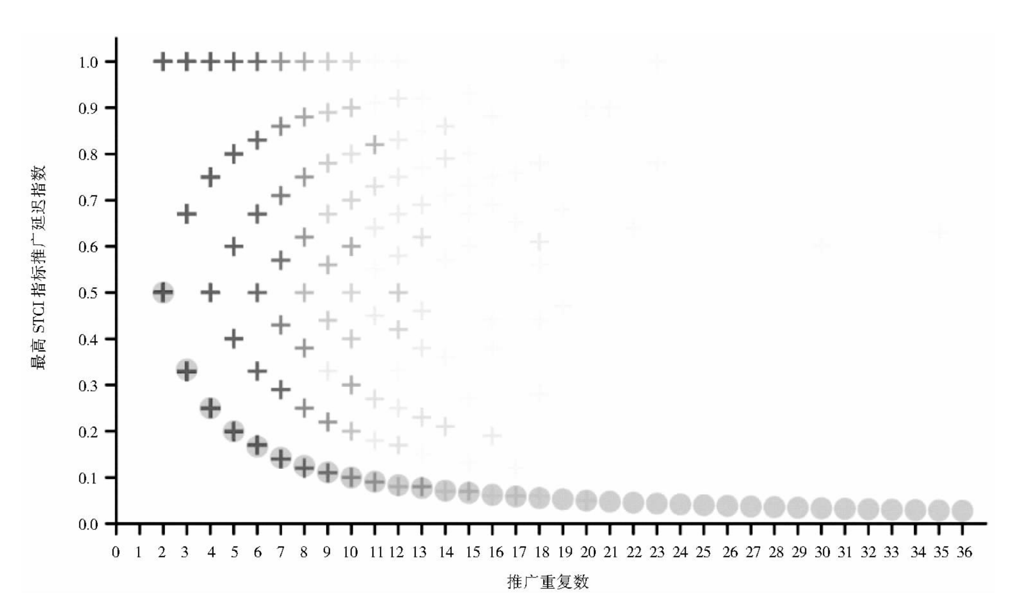


图 4　最高 STCI 推文延迟指数分布

3.5　高 STCI 推文内容分析

将 56 355 条微信公众号推文标题先去除标点符号，利用"jieba"中文分词库提取标题里具有实际意义的名词(n)、动词(v)、动名词(vn)，共 653 779 个，去重后有 35 036 个。选取频率超过 200 并去掉不表示实际意义的词后获得高频分词 80 个，表 5 为前 80 位推文标题高频词统计。将这个 80 * 80 的共词矩阵导入 Ucinet 软件并用 NetDraw 可视化工具绘制推文标题高频分词共现网络图谱，见图 5。

表 5　前 80 位推文标题高频词统计

高频词	频率	词性	高频词	频率	词性	高频词	频率	词性	高频词	频率	词性
女人	3 222	n	思考	482	v	老婆	279	n	感动	227	v
孩子	2 123	n	值得	463	vn	人品	278	n	教育	226	n
男人	1 812	n	女儿	435	n	儿子	270	n	简单	226	n
自己	1 667	n	生命	407	n	优秀	270	n	真实	222	n
人生	1 616	n	读书	399	n	聪明	268	n	高考	222	n
中国	1 415	n	结婚	387	v	女生	265	n	暴露	221	v
生活	1 264	n	微信	384	n	出轨	265	vn	文学	219	n
故事	1 061	n	懂得	372	vn	历史	261	n	爸爸	218	n
婚姻	1 003	n	内涵	365	n	岁月	259	n	关系	215	n
世界	847	n	老公	349	n	努力	249	n	成熟	214	n
朋友圈	821	n	家庭	343	n	父亲	245	n	层次	210	n
漫画	808	n	朗诵	339	vn	收藏	245	v	遇见	210	v
经典	789	n	教养	322	n	修养	241	n	老师	210	n
幸福	693	n	真相	306	n	三观	241	n	早起	210	vn
父母	609	n	离婚	305	n	未来	241	n	漂亮	206	n
朋友	577	n	照片	305	n	母亲	237	n	快乐	206	n
妈妈	529	n	说话	300	v	时间	236	n	灵魂	205	n
决定	513	vn	善良	292	n	幽默	234	n	做人	204	vn
爱情	509	n	情商	291	n	家长	231	n	诗词	203	n
最美	486	n	一辈子	290	n	夫妻	227	n	细节	201	n

这是一个网络密度为 3.01 的全联通网络。图 5 中圆点表示高频分词，圆点的大小表示该分词在推文标题出现的频率，圆点越大，出现频率越大。圆点之间连线粗细表示高频分词之间的共现频率的大小。通过

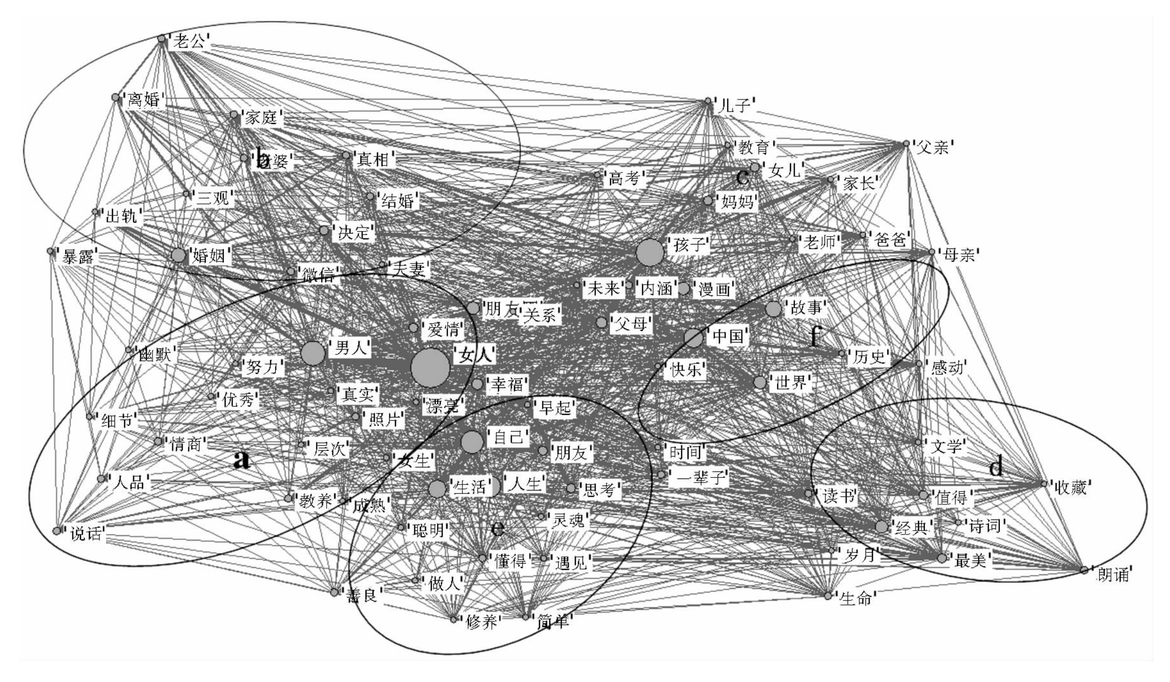


图5　推文标题高频分词共现网络图谱

聚类分析，将微信公众号推文涉及到的主要内容进行分类：a、男人与女人之间的爱情、关系，有关优秀、人品、情商、努力、幽默的男人相关话题，有关漂亮、照片、幸福、成熟的女人相关话题；b、婚姻中老公或老婆出轨、离婚的家庭问题；c、与家长和老师有关的孩子教育问题，孩子高考及未来发展；d、经典文学、诗词的收藏、阅读、朗诵；e、关于人生、生活的思考，自己与朋友之间的关系，做人的道理；f、中国与世界的新闻、历史故事。

3.6　高 STCI 推文内容分析

为了进一步得到微信推文内容特征，本文利用了人工逐篇阅读高 STCI 推文（STCI 排名前 30 位）和低 STCI 推文（STCI 排名后 30 位）的方法。表6为 STCI 排名前10位的推文统计表，表7为 STCI 排名后10位的推文统计表。

表6　STCI 排名前10位微信推文统计

STCI	公众号	题名
1 316.74	有书	明星恋爱致网络瘫痪，而这个外国集体跪求的中国男人却无人问津……
1 311.84	十点读书	有格局的人，绝不会在这些烂事上纠缠
1 310.6	十点读书	给出轨老公的一封信
1 303.73	有书	我很惭愧，以这种方式认识您：他用58年助黄河水变清，却到去世才成新闻
1 280.19	有书	她中年丧夫，51岁入狱，历经半生磨难，90岁却活成中国最精致的女人！
1 266.35	有书	有教养的人，从不做这4件事
1 261.84	有书	给出轨老公的一封信
1 259.17	十点读书	在朋友圈里，千万别抱怨
1 255.64	十点读书	好的婚姻，一定要谈钱
1 253.69	十点读书	收到请回复

逐篇阅读后可知，发表高 STCI 推文较多的微信公众号主要有十点读书、有书，分别是19篇和11篇，发表低 STCI 推文较多的微信公众号主要有美文共享、智慧管理，分别是18篇、7篇。结合前文聚类分析的结果来看，这30篇高 STCI 推文主题并没有涵盖经典文学、诗词的收藏、阅读、朗诵这几方面。这说明，当前社会正面临着“深阅读危机”。在数字时代，互联网精确地释放出各种类型的感官刺激和认知刺激（反复、瞬时、高强度、智能、交互式）。并且，很多信息都唾手可得，毫不费力。同时，面对纷繁的信息数据人们常常高度紧张且一心多用。在外在和内在因素的共同作用下读者阅读的速度更容易增加，阅读内容更容易快速切换，注意力更容易分散，其代价则是理解力的降低。相反，时间便成为了稀缺的资源，哪怕是短暂的思考，也可能是浪费时间，“深阅读”离人们的生活渐行渐远。

通过对30篇低 STCI 推文研读后发现这些推文的内容都涉及广告推销。如“年销10亿盒，畅销40年，欧洲传奇”国民美味“到底凭什么?”“G20指定毛巾，改变一种生活方式，4条99元!”。这些内容消耗了读者对该公众号的信任，读者一看到标题就会选择性的跳过。

进一步，对30篇高 STCI 推文内容分析后得出高 STCI 推文特征主要包括：

（1）满足读者的好奇心。①设置带有悬念及争议性的标题。如推文“明星恋爱致网络瘫痪，而这个外国集体跪求的中国男人却无人问津……”用明星恋爱的

王磊，江仔玲，刘宇婷.社会化阅读环境下阅读内容计量研究——以阅读推广类微信公众号推送文章为例[J].图书情报工作,2019,63(2):12-23.

表7 STCI排名后10位微信推文统计

STCI	公众号	题名
101.8	时代书屋	让孩子成为古埃及小专家的考古手册——神秘日志·精华版:木乃伊的诅咒
114.48	美文共享	最适合亚洲宝宝的乳粉,值得试试
116.74	美文共享	G20指定毛巾,改变一种生活方式,4条99元!
120.13	管理智慧	2017年最超值的学习社区,让企业转型不再难!(机会难得)
124.68	美文共享	夏日精美好物,美文共享特别推荐
131.07	管理智慧	如何转型"新零售",让业绩增长50%?(小程序上线,体验升级)
133.42	美文共享	夏日精美好物,美文共享特别推荐
134.93	管理智慧	2017年最超值的学习社区,让企业转型不再难!(机会难得)
137.21	美文共享	精美好物等你来挑
137.59	管理智慧	2017年最有价值的学习社区,华为、京东、苹果的管理者都加入了!

高热度与文章主人公的无人问津这一强烈对比设置悬念(这个有如此大魅力的中国男人到底是谁),读者为一探究竟就会点击阅读。推文"好女人,一定要学会谈钱",在标题处就引发争议(金钱在爱情和婚姻中所占的比重大小),制造话题,读者看到标题后便会忍不住想知道推文是如何阐述这一话题的。②设置密集的点阵式刺激点。推文"一等奖作品:《我娘是个疯子》,令无数人泪奔……"开篇制造悬念(娘为什么是疯子),而后衔接"疯娘"的种种悲惨遭遇、"我"与"疯娘"的种种冲突,最后矛盾化解之时"疯娘"又意外命丧悬崖。故事情节环环相扣又兼具反转,几乎每两屏出现一个刺激点,让人欲罢不能。

(2)满足读者的自我表达需要。①设置能引起人共鸣的标题。推文"那些拿命换钱的年轻人,换着换着就死了"标题直戳人心,引起读者的强烈共鸣,引发阅读兴趣。"千万不要小瞧一个30岁的女人",其标题直接道出了30岁女性的心声,虽然她们已不再是花季少女,但随着年龄的增长,生活的历练,她们已然成为了女战士,可以坦然面对来自生活的种种考验。②从熟悉的生活场景出发为读者解决问题,并帮助读者表达出他们内心的想法。推文"中国妈妈为什么怨气重?婆婆越位、丈夫缺位、自己错位"从经典的家庭矛盾出发,并分析原因是婆婆越位、丈夫缺位、自己错位,得到读者的高度认可,并提出切实可行的解决方法——各就各位,符合读者需求。推文"永远别和层次低的人纠缠"选取生活中熟悉的场景:每个人都遇到过蛮不讲理的人,跟他们讲道理往往会惹上一身麻烦,得不偿失。熟悉的情境让读者身临其境的同时,还提出了解决办法就是不理会。

(3)满足"屏阅读"时代特征的写作法则。①语言简单化、口语化,能用短句就不用长句。如推文"某岛国在南海边花3年造的珊瑚岛离奇消失,竟说被中国海星吃了,中国:别瞎说啊"行文中多使用短句,文章富有节奏感,音律感,降低了读者的理解成本。②用文字营造画面感。推文"一等奖作品:《我娘是个疯子》,令无数人泪奔……"中利用众多细节描写栩栩如生地刻画了"疯娘"这一神志不清却深爱着孩子的母亲形象,震撼读者的内心。③使用"金句"画龙点睛。推文"好女人,一定要学会谈钱"在三个故事中穿插诸如"等你有了足够的钱,才能有底气,有资格去追逐你觉得真正重要的东西。"等精辟语句,给读者带来强烈的冲击感。④用具体代替抽象。推文"给出轨老公的一封信"中描写主人公刚发现老公出轨"这一行行字眼,个个都像陨石那么大,毫无预兆又雷霆万钧地砸过来,砸得我五脏六腑都碎成了渣。"将老公与其他女人的短信比喻成陨石砸在心上,一个伤心欲绝的女人形象瞬间浮现在读者眼前。⑤表达克制。推文"她中年丧夫,51岁入狱,历经半生磨难,90岁却活成中国最精致的女人!"中的主人公郑念一生坎坷,每个时期都值得大肆书写,但作者用词精辟、贴切,跌宕起伏的经历结束地干脆、果断,碎片化的描写让人意犹未尽。⑥舒适的排版。阅读推广类推文篇幅通常较长,许多高STCI推文采用多段少句型排版,并通过字体加粗、下划线、添加字体背景色、改变字体颜色等方式凸显重点内容。

4 结论与建议

4.1 结论

本文利用社交媒体大数据平台和文本挖掘技术对微信公众号(阅读推广类)基本信息和公众号推文内、外部特征数据进行采集,之后,运用统计学与信息计量学相关理论及技术对采集到的数据进行实证研究,如下:

(1)微信公众号的推文数量与其阅读数、平均阅读数、点赞数、平均点赞数不存在明显的相关关系,但其平均阅读数与平均点赞数之间存在较强的正相关关系,即平均阅读数越多的公众号往往平均点赞数越高。

(2)基于分类、认证、地域分布的公众号统计,发现地处沿海及华中经济发达地带的微信公众号的WCI指数相对其他地区更高,即地区经济发展水平会影响该地区居民的阅读方式,经济发达地带的居民较其他地区的居民更习惯于在微信平台进行阅读。同时还发

现认证的公众号更具有公信力，即用户更倾向于选择获得认证的微信公众号推文进行阅读。

（3）由基于重复推文次数的各公众号耦合分析发现微信公众号之间互相借鉴推文是普遍现象，同时，拥有较多相同推文的公众号之间传播能力不尽相同。

（4）由最高 STCI 推文延迟指数分析发现推文的时效性和重复性与该推文的传播指数无关，即用户阅读推文时并不关心其是否为原创。

（5）由高 STCI 推文内容分析本文发现传播能力强的微信推文主要涉及男女关系、儿童教育、成功等生活性内容，而较少涉及朗诵、诗词等文学性内容，即新媒体环境下社会正面临着"深阅读危机"。

（6）对低 STCI 推文研读后本文发现这些推文的内容大都涉及广告推销。对高 STCI 推文内容分析后本文得出高 STCI 推文主要特征包括：①设置带有悬念、争议性的标题；②设置密集的点阵式刺激点；③设置引人共鸣的标题；④从熟悉的生活场景出发为读者解决问题；⑤文本语言简单化、口语化；⑥用文字营造画面感；⑦使用"金句"画龙点睛；⑧用具体代替抽象；⑨表达克制；⑩排版舒适。

4.2 建议

基于研究结论，提出了如何提高阅读推广类微信公众号运营能力的一些建议。

（1）合理利用微信后台数据（订阅量、每日独立 IP 访问量等指标），并结合微信大数据平台的各项指标（如点赞数、阅读数、WCI 指数）或自定义指标（如本文的 STCI 或 HSPDI 指数）找到用户的兴趣点，精准推送。

（2）加大微信公众号的宣传力度，深入经济较不发达地区，拓宽受众面。并积极进行认证工作，优化微信菜单栏，提高用户体验度。

（3）针对目前阅读推广类微信公众号之间较为严重的推文借鉴现象，建议各公众号加强对自身原创推文知识产权的保护，这有利于树立公众号的品牌。与其转，不如造，打造高传播指数的原创推文应该是各微信公众号努力的方向。

（4）对高 STCI 推文特征总结后，本文认为创作一篇高传播指数的微信推文需要满足"屏阅读"时代特征的写作法则，同时，满足读者的好奇心或自我表达需要。

致谢：感谢孙菁、易永梅在论文写作中数据描述、文字校对和中文摘要翻译所提供的帮助。

参考文献：

[1] 赵文军，陈焕之，蒋伟进. 近 10 年来国内外数字阅读研究综述[J]. 图书情报工作，2017，61(18)：128－136.

[2] 张晓梅. 浅议阅读推广中的服务创新[J]. 图书情报，2017，61(1)：29－30.

[3] 何韵，何兰满. 从传统阅读与数字阅读的二元关系论全民阅读推广策略——以日本为例[J]. 图书馆，2015(7)：34－38.

[4] ANTONIO J，ALONSO J，GOMEZ R，et al. Social reading：platforms，applications，clouds and tags[M]. Oxford：chandos pulishing ltd，2013.

[5] 孙一钢，张炜，谢强，等. 国家数字图书馆的数字阅读服务与推广[J]. 图书馆杂志，2016(10)：4－8.

[6] 张晓青，张植禾，相春艳. 基于 Web2.0 的知识传播研究[J]. 现代传播（中国传媒大学学报），2010(4)：123－126.

[7] 姜志芳. 信息用户需求及信息用户行为研究综述[J]. 图书馆论坛，2006(4)：284－286，278.

[8] 杨婧. 基于社交媒体的图书馆阅读推广研究[J]. 图书馆研究，2017(5)：57－60.

[9] LI W，WU Y. Adolescents' social reading：motivation，behaviour，and their relationship[J]. Electronic library，2017，35(2)：246－262.

[10] 张云，茆意宏. 高校学生社会化阅读行为调查分析[J]. 国家图书馆学刊，2016，25(2)：77－87.

[11] 刘静. 我国高校图书馆认证用户微博调查分析——以新浪微博为平台[J]. 图书馆学研究，2012(1)：90－95.

[12] KIBUGI S. The use of social media in the dissemination of information in selected public and private university libraries in Kenya [J]. Innovation：journal of appropriate librarianship and information work in Southern Africa，2013(47)：100－120.

[13] ZINOVIEV D，DUONG V，ZHANG H. A game theoretical approach to modeling information dissemination in social networks [EB/OL]. [2018－10－10]. https://arxiv.org/pdf/1006.5493v1.pdf.

[14] YANG D，CHOW T W S，ZHONG L，et al. True and fake information spreading over the Facebook[J]. Physica a-statistical mechanics & its applications，2018，505：984－994.

[15] CHANCE R，LESESNE T. Rethinking reading promotion[J]. Teacher librarian，2012，39(5)：26－32.

[16] DIKE V W，EKE H N，BABARINNDE E T. Social media and reading among secondary school students in Enugu State[J]. Mousaion，2013(31)：60－81.

[17] SAMPENE A B，AFORO A A. Impact of social media on academic reading：a study at Kwame Nkrumah University of Science and Technology，Kumasi，Ghana [J]. Asian journal of humanities and social studies，2014，2(1)：92－99.

[18] LI W，WU Y，LIU Y. Study on relationship between adolescents' social reading motivation and their behaviors：a case study of middle school students' reading via Wechat in Shanghai [J]. Library & information service，2014，40(6)：115－128.

[19] SUN L，XIE J. An exploration of reading promotion and readers' advisory at Chinese public libraries[J]. Reference & user services quarterly，2016，56(2)：96.

[20] 苑世芬. 高校图书馆新媒体阅读推广策略研究[J]. 现代情报,2013,33(1):74－77,81.

[21] 张秋，杨玲，王曼. 高校图书馆微信公众平台服务发展现状及对策[J]. 图书馆建设,2014(2):61－65,69.

[22] 万慕晨，欧亮. 基于微信公众平台的高校图书馆阅读推广效果实证研究[J]. 图书情报工作,2015,59(22):72－78.

[23] 李武，吴月华，刘宇. 青少年社会化阅读动机与行为之关系研究——以上海市初高中生微信阅读为例[J]. 图书情报工作,2014,58(23):61－68.

[24] 清博大数据. 清博大数据平台[EB/OL]. [2017－12－19]. http://www.gsdata.cn/.

[25] 清博指数. 微信传播指数 WCI(V 13.0)[EB/OL]. [2017－12－19]. http://www.gsdata.cn/site/usage.

[26] 王磊，吕鹏辉，张仁琼. 高校图书馆微信信息计量分析[J]. 图书情报知识,2017 (5):49－58,88.

作者贡献说明：

王磊：负责主题与大纲制定、数据处理及分析、数据汇总、论文修改；

江仔玲：负责论文撰写及修改、数据处理及分析；

刘宇婷：负责数据处理及分析、数据汇总。

Informetrics Study of the Reading Contents in Social Reading Environment: Take the Articles Published by Wechat Public Platform of Reading Promotion for Example

Wang Lei[1] Jiang Ziling[2] Liu Yuting[3]

[1] Hefei University of Technology Library, Hefei 230009

[2] School of Instrument Science and Opto－electric Engineering, Hefei University of Technology, Hefei 230009

[3] Textile and Clothing of Engineering, Soochow University, Suzhou 215021

Abstract: [**Purpose/significance**] With the development of the social media big data platform, evolutionary of the big data and database technology, the various data are becoming available and accessible in web 2.0. In this context, the paper tries to explore the feasibility of informetrics study of reading contents based on the big data of social media. [**Method/process**] The paper applies the big data platform and text mining technology to collect the basic information of WeChat public account and the contents of the published articles. Then, the collected data are counted. The distribution of WeChat public account on classification, geographical and certification, as well as the number of published articles, reading number, average reading number, thumbs up number and average thumb up number up of WeChat public account are obtained. In addition, the Single Tweets Communication Index (STCI), the high frequency keywords in the title co-occurrence matrix and coupling matrix based on repetitions of WeChat public account articles are proposed by using the theories and techniques of the informetrics. At the same time, the paper puts forward the Publish Delay Index of Highest STCI Article (HSPDI) algorithm to investigate the effect of repeating publish times of articles on the communication effect of the articles. Finally, the paper summarizes the content characteristics of articles through artificial reading of high STCI articles. [**Result/conclusion**] There is a positive correlation between the average reading number and the average thumb up number in Wechat public accounts. Many wechat public accounts have the same phenomenon of publishing the same article. Whether the first time to publish the article has no obvious effect on its STCI. Furthermore, after analyzing the content of high STCI tweets, the paper puts forward three main features of the high STCI articles, including satisfaction of the readers' curiosity, the meeting of the reader's self-expression needs, and the compliance of the writing principle of "screen reading".

Keywords: social reading social media WeChat Public account informetrics